Eur. Phys. J. E **31**, 411–417 (2010)
DOI 10.1140/epje/i2010-10590-6

THE EUROPEAN
PHYSICAL JOURNAL E

Regular Article

Free-energy-based method for step size detection of processive molecular motors

B. Bozorgui[1,2], K. Shundyak[2], S.J. Cox[3], and D. Frenkel[3,2,a]

[1] Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, USA
[2] FOM Institute for Atomic and Molecular Physics, Science Park 104, 1098 XG Amsterdam, The Netherlands
[3] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

Received 11 December 2009 and Received in final form 18 February 2010
Published online: 20 April 2010 – © EDP Sciences / Società Italiana di Fisica / Springer-Verlag 2010

Abstract. We report a free-energy-based algorithm to estimate the step size of processive molecular motors from noisy, experimental time position traces. In our approach, the problem of estimating step sizes reduces to the evaluation of the free energy of directed lattice polymers in a random potential. The present approach is Bayesian in spirit as we do not aim to determine the most likely underlying time trace but rather to determine the step size and stepping frequency that are most likely to yield the observed data. We test this method on synthetic data for the simple case of noisy traces with fixed underlying step size and Poissonian stepping statistics. We find that the present scheme can work at signal-to-noise levels that are about 40% worse than those where the best existing step detection methods fail. More importantly, the present approach yields a much more accurate estimate of the step size. Although we focus on the case of non-reversing walks with a single step size, we show that we can detect if this assumption is violated. In principle, the method can be extended to more complex stepping scenarios but we find that for noisy data, multi-parameter fits are not reliable.

Introduction

Information about the step size and stepping frequency of processive molecular motors can be obtained from optical experiments that follow the motion of an optically visible "tag" attached to the motor [1–4] (for a recent review, see, *e.g.* [5]). In these experiments, the spatial resolution with which steps can be resolved is usually limited by the statistical noise in the data. For this reason, sophisticated data analysis is required to extract the underlying sequence of fixed-amplitude steps from the noisy measurements. As can be seen from the example in fig. 1, direct inspection of the noisy data is often not good enough to reconstruct the underlying motor motion. The figure shows a synthetic time trace that is the sum of a specific stepping sequence with Poisson-distributed fixed steps of size Δ_r (the subscript r stands for "real") and uncorrelated Gaussian noise with zero mean and standard deviation σ. The outcome of the fitting procedure will be an estimated step size Δ. More precisely: a reliable step detection algorithm should find that the assumption $\Delta \approx \Delta_r$ describes the experimental data significantly better than any other value of Δ. Existing methods to extract the most likely value of Δ_r from the noisy data set (see ref. [5]) approach this problem by attempting to decompose the observed data into pure noise and a unique, underlying "noise-free" stepping curve. Several methods exist to estimate this underlying time trace. Examples are the step function fitting of ref. [6], the Schwartz information criterion [7], velocity thresholding [8], or "filtering" methods [5]. Statistical analysis of these fits is then used to obtain information about step size(s) and stepping rate(s). Of course, the performance of these schemes depends strongly on the robustness of the fitting procedure [5], and the robustness of all these methods drops sharply when the data set is so noisy that many stepping histories (indicated for instance by the two blue traces in fig. 1) fit the experimental data equally well. However, as we shall argue below, the objective of experiments is primarily to determine the parameters (step size, step frequency) that characterize the motion of the motor and it is usually irrelevant which one of the many possible fits is the "true" noise-free time trace. Hence in our approach all possible traces are considered and the optimal stepping parameters are obtained by considering the "partition function" of this collection of traces. In what follows, we describe our approach and apply it to synthetic data, because in that case we know the exact answer that our data anaylsis should yield if it works well.

[a] e-mail: `df246@cam.ac.uk`

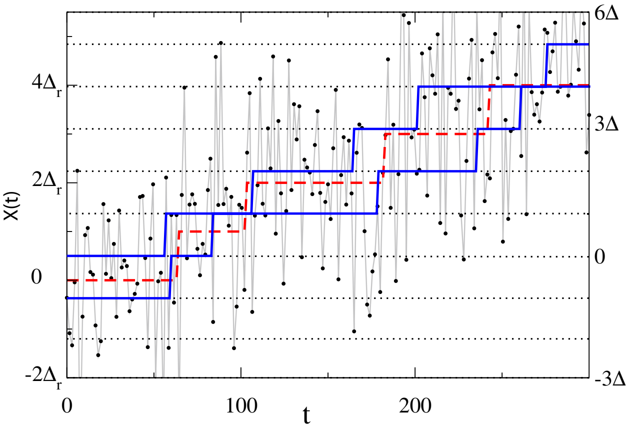

Fig. 1. (Color online) Synthetic data set of a position-time trace $X(t)$ (●, black). The "time" t labels the consecutive data points. The figure shows an underlying stepping trajectory with step size Δ_r (dashed, red) and added Gaussian noise with amplitude σ. For the example shown, the noise amplitude $\sigma = 1.0\Delta_r$. The two blue traces correspond to two distinct trial step sequences with an incorrect step size $\Delta = 1.2\Delta_r$ and a constant offset Δ_0. The left-hand ordinate measures the displacement $X(t)$ in units of Δ_r. The-right hand ordinate shows the displacement in units of Δ, relative to the offset Δ_0.

Methods

To explain the approach that we follow, we focus on the simple example of motors that move processively with steps of fixed size Δ_r. In what follows, we make the crucial assumption that the underlying stepping process is Markovian. Whilst this assumption is probably justified in some cases (*e.g.*, for the processive motor protein kinesin-1 moving along a single microtubule fixed on a substrate [1, 9]), there are also examples where the assumption of a single, Markovian stepping process has been shown to be an over-simplification [10].

In what follows, we use synthetic "experimental" data that mimic a series of measured displacements $x_1, x_2, x_3, \cdots, x_t, \cdots, x_N \equiv \{\vec{x}\}$, where the index t denotes the (discretized) time.

Our test data are generated as the sum of a "hidden" directed random walks of step size Δ_r and of uncorrelated Gaussian noise of zero mean and constant standard deviation σ. We assume that the noise level σ in real experiments can be estimated from a part of the time trace where the motor is not moving. In what follows, we shall vary the ratio σ/Δ_r. The data sets consist of $N = 4 \times 10^3$ points, corresponding to, on average, 40 steps. We use this relatively small number of steps because most experiments are limited to a similarly small number. Longer runs would allow us to detect steps in systems with lower signal-to-noise ratios. If we use Δt to denote the (fixed) time interval between successive data points, then the total duration of measurement corresponds to $N\Delta t$. As we show towards the end of the present paper, it is possible to generalize the present method to walks that involve more than one step size or to walks that include backstepping. However, initially we focus on the simplest case.

There are many possible underlying walks (denoted by $y_1, y_2, y_3, \cdots, y_N \equiv \{\vec{y}\}$) that are compatible with the (noisy) experimental data $\{\vec{x}\}$, but some fit the data better than others. In the case where successive jumps are Poisson distributed, the underlying walk can be viewed as a hidden Markov process. We define the statistical weight of a specific sequence of steps $\{\vec{y}\}$, given the data set $\{\vec{x}\}$ as

$$w(\vec{y};\vec{x}) = \prod_{t=1}^{N} \exp\left[-\frac{(x_t - y_t)^2}{2\sigma^2}\right]. \tag{1}$$

In Bayesian parameter estimation, the objective is to obtain the best possible estimate of the underlying parameters to characterize the stochastic process at hand, in the light of the available data (see, *e.g.* [11]). To this end, we need two ingredients: 1) an expression for the probability to observe the data sequence, given the parameters and 2) an expression for the prior distribution of the parameters. The most straightforward approach to Bayesian parameter estimation would be to determine the underlying stepping sequence that is most likely, given the data (as in eq. (1)). However, in our approach, we sum over all distinct stepping sequences and only determine the step size (Δ) and step probability (p) that are most likely, given the data. To this end, we introduce a "partition function" $Z_{\{\vec{x}\}}(\Delta)$ that is the sum of all possible terms of the type given in eq. (1) for walks with jump size Δ and jumping probability p,

$$Z_{\{\vec{x}\}}(\Delta) \equiv \sum_{\text{paths}} \exp\left[-\sum_t \frac{(x_t - y_t(\Delta))^2}{2\sigma^2}\right] p^n (1-p)^{N-n}, \tag{2}$$

where n denotes the number of steps in a given trajectory. In what follows we assume that the prior distributions of Δ and p are uniform in the interval $\{0, x_{\max}\}$ and $\{0, 1\}$, respectively, where $x_{\max}$ is the total displacement of the motor during the experiment. As we have now summed over all paths, the probability of a given set Δ, p is proportional to the partition function Z in eq. (2). We stress that p is not known *a priori* and should therefore be obtained by finding the value that maximizes Z. However, in the present example most paths that contribute to the partition function contain very nearly the same number of steps and then p can simply be chosen equal to n_{steps}/N.

It is straightforward to compute Z numerically. At first sight, such a calculation may seem to be a daunting task because the number of distinct trajectories increases exponentially with time. However, as the traces behave as non-reversing polymers, their partition sum can be computed recursively in a time that scales linearly with the length of the time interval. To this end, we use the "moment propagation" algorithm described in refs. [12, 13]. This algorithm is an exact transfer matrix scheme that is well suited to compute the partition function of non-self-avoiding lattice polymers in an arbitrary potential. In our calculations, we assume that the *a priori* probability of a walk with n steps is $p^n(1-p)^{N-n}$. Of course, many of these traces will be far removed from the data points and will therefore have a very low statistical weight. In practice, this means that effectively all trajectories with an appreciable weight are confined to a "tube" of width 5σ around the (unknown) true trajectory. This feature greatly reduces the computational effort. A second point to note is that a family of trial time traces is not just characterized by the step size Δ, but also by the starting position Δ_0. In many experiments, a good estimate of Δ_0 can be obtained from a part of the time trace where the motor is not moving. However, even if Δ_0 is not known, it can be obtained quite easily by computing the partition sum for different starting points and selecting that value of Δ_0 that yields the largest value of Z. In what follows, we shall initially consider the case that $\Delta_0 = 0$ (because this simplifies the notation). Subsequently, we show that crucial information can be obtained by considering the variation of Z with Δ_0.

Results and discussion

By analogy to the statistical physics of polymers, we can define a free-energy density $f(\Delta)$ as $f_{\{\vec{x}\}}(\Delta) = -\frac{1}{N}\ln Z_{\{\vec{x}\}}(\Delta)$. The subscript $\{\vec{x}\}$ indicates that the value of the free-energy density depends on the experimental data points $\{\vec{x}\}$. Clearly, the value of $f_{\{\vec{x}\}}(\Delta)$ depends on the value of the trial step size in the walks $\{\vec{y}\}$. In the case that the data have very little noise, the lowest free-energy path will be the one for which Δ is equal to the true step size Δ_r. However, in general the free energy alone is not a reliable indicator of the quality of the fit: for higher noise amplitudes $\Delta = \Delta_r$ need not correspond to the absolute minimum in the free energy. Figure 2 shows the free energy $f_{\{\vec{x}\}}$ as a function of Δ/Δ_r for three different values of σ. We vary Δ in steps of size δ, such that $\delta/\Delta_r = 2.5 \cdot 10^{-3}$. For the lowest-noise data, a sharp minimum in the free energy, corresponding to the correct step size Δ_r is clearly visible in fig. 2. As is to be expected, the minimum becomes less pronounced as the noise level increases. We observe that secondary (local) minima can sometimes also be seen at $\Delta = \Delta_r/n$, with n an integer. This is not surprising as a trajectory with step size Δ can be retraced fairly well by a trajectory with steps of $\Delta/2$. However, although we usually find free-energy minima at fractions of the true step size, we never find significant minima at multiples of the true step size. Hence, the free-energy minimum with the largest value of Δ singles out the prime candidate for the true step size. The key question to address is whether the observed features in the free energy are statistically significant. We address this question in two stages: first by comparing the computed free energy with a "reference" free energy and, subsequently, by studying the dependence of the free energy on the value of Δ_0.

To account for the trivial dependence of the computed free energy on the assumed step size Δ, we introduce a reference free energy $f_{\text{ref}}(\Delta)$ defined as the average free energy that would have been found if the experimental x-t trace would have had a step size Δ, rather than Δ_r, but the same value of the noise amplitude σ and the same overall displacement. We can now introduce a free-energy difference $\delta f(\Delta)$ defined as

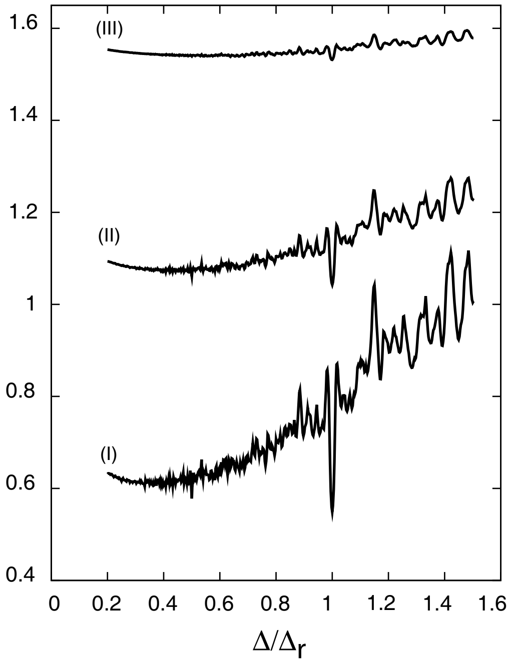

Fig. 2. Free energy f as a function of the dimensionless step size Δ/Δ_r for three different noise levels $\sigma/\Delta_r \approx 0.4$ (I), 0.68 (II) and 1.1 (III). For clarity, the curves II and III are shifted by 0.25 in the vertical direction. Note that, in addition to peaks at the correct step size Δ_r, subsidiary peaks at $\Delta_r/2$ are also visible.

$$\delta f(\Delta) \equiv f_{\{\vec{x}\}}(\Delta) - \bar{f}_{\text{ref}}(\Delta), \tag{3}$$

where $\bar{f}_{\text{ref}}(\Delta)$ is the average reference free energy. We compute this average by evaluating f_{ref} for many different data sets with same step size and noise level.

This free-energy difference should be close to zero when $\Delta \approx \Delta_r$. In fact, fig. 3 shows that $\delta f(\Delta)$ has a sharp minimum at $\Delta = \Delta_r$, even though the noise amplitude is equal to $0.6\Delta_r$. We chose this specific value of the signal-to-noise ratio because, for the data set used, existing fitting methods fail if $\sigma \geq 0.6\Delta_r$ (see below).

In order to test whether $\delta f(\Delta)$ differs significantly from zero, we must estimate the statistical noise in this quantity: σ_f. This we do by generating many $f_{\text{ref}}(\Delta)$ curves: the average yields $\bar{f}_{\text{ref}}(\Delta)$ and the root-mean-square variance provides a measure for the expected statistical noise in $f_{\text{ref}}(\Delta)$. A first criterion in the step selection process is the following: a step size Δ is considered as a possible candidate when the value of $\delta f(\Delta)$ is within a distance of order $\mathcal{O}(\sigma_f)$ from the line $\delta f = 0$. As can be seen for fig. 3, the primary peak $\Delta/\Delta_r = 1$ is significant at $\sigma/\Delta_r = 0.6$. However, for higher noise levels ($\sigma/\Delta_r = 1.05$), $\delta f(\Delta)$ does not differ significantly from zero over a wide range of Δ values. This would suggest that we cannot determine the true step size in this high-noise regime. In fact, the situation is not as bad as it seems, as it turns out that the statistical noise in $\delta f(\Delta)$ is strongly

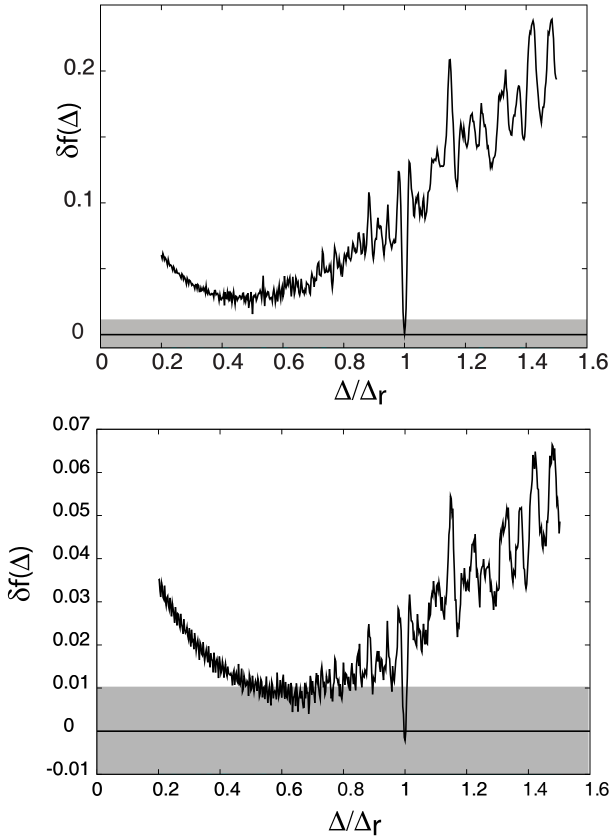

Fig. 3. Free-energy difference $\delta f = f(\Delta) - \bar{f}_{\rm ref}(\Delta)$ (see text) as a function of the dimensionless step size Δ/Δ_r for $\sigma/\Delta_r = 0.6$ (top curve) and $\sigma/\Delta_r = 1.05$ (bottom curve). The average reference free energy $\bar{f}_{\rm ref}(\Delta)$ is obtained by averaging of 1000 $f_{\rm ref}$ curves. The grey band around $\delta f = 0$ indicates the estimate error in δf.

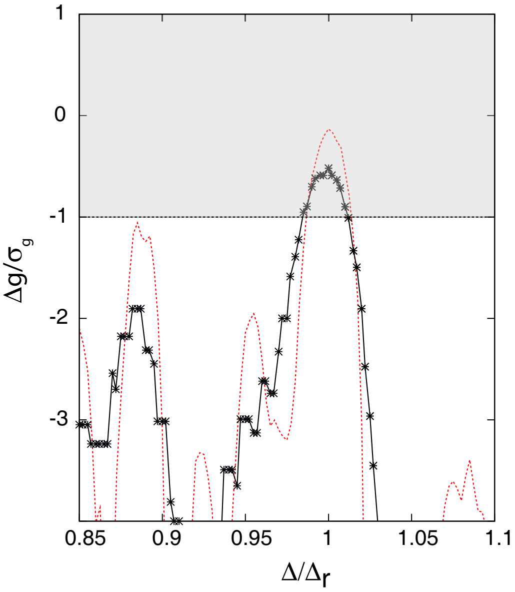

Fig. 4. (Color online) The amplitude of the modulation Δg (see text) of the free energy with offset Δ_0. As explained in the text, we compute the difference in modulation amplitude between the free-energy modulation computed for the data set, and the modulation of the reference free energy. The figure shows results for $\sigma/\Delta_r = 1.05$. In other words: the noise is slightly larger than the step size. The dashed curve shows $\Delta g/\sigma_g$ (see text) in the region of the primary minimum of the free energy. In addition to the true maximum at $\Delta_r = 1$, there is a spurious peak at $\Delta/\Delta_r = 0.88$. By filtering the data (see text), the amplitude of the spurious peak is suppressed such that it lies at the boundary of the 95% confidence interval. The gray band denotes the $\pm\sigma_g$ error margin.

correlated. Hence, features such as the peak at $\Delta = \Delta_r$ may still be significant, even if the overall level of δf is not significantly different from zero over a wide range of Δ values. It is at this stage that we consider again the dependence of the free energy on the offset Δ_0. We should expect that if we choose the correct value of Δ_0 (say, $\Delta_0 = 0$), then the free energy will be at a local minimum, whereas if we choose an offset shifted by $\Delta_r/2$, then we should expect to have a high free energy (as the test trajectories are always $\Delta_r/2$ removed from a possible true trajectory). As $f(\Delta, \Delta_0)$ is a periodic function of Δ_0 with period Δ, the method to find the correct Δ_0 is to compute the variation of $f(\Delta, \Delta_0)$ over one period. In addition to Δ_0, this calculation also yields the amplitude of the modulation of f with Δ_0. We denote this amplitude by Δf (not to be confused with δf that measures the distance to the reference free energy). As before, we also compute the modulation of the reference free energy where the underlying step size is equal to Δ rather than Δ_r. In this way, we obtain $\Delta f_{\rm ref}$, the average modulation of the reference free energy with Δ_0, and σ', the estimated statistical error in this quantity. As a last step, we now compute $(\Delta f - \Delta f_{\rm ref})$. We denote this quantity Δg. In the first step of the analysis, we have identified possible regions where δf does not differ significantly from zero. We now focus on these regions only and we ignore the regions near “sub-harmonic” peaks in δf at fractional step sizes. In the region of interest, we compute $\Delta g/\sigma_g$. We expect this quantity to be close to zero only when $\Delta \approx \Delta_r$. Figure 4 shows that this is indeed the case. In this figure, we do not show the results for the (irrelevant) sub-harmonics.

Using this approach, we can easily obtain a good estimate of the true step size under conditions where the noise is less than 80% of the step size. The best performance is achieved if we take into account that the peak in Δg has a finite width. This means that for a true peak, neighboring points are correlated, but this is not the case for peaks that are due to statistical noise. This fact allows us to obtain a slightly better estimate of the peak amplitude by convoluting the Δf and $\Delta f_{\rm ref}$ with a filter function that has (approximately) the same shape as the peak in Δf. Using the last refinement, we are able to obtain reliable estimates for the step size Δ under conditions where the step size is comparabe to the noise $\sigma/\Delta_r = 1.05$. The method runs out of steam when $\sigma/\Delta_r \geq 1$. These numbers are significant because a recent review of existing step fitting methods has concluded that, on balance, the so-called χ^2-method of ref. [6] outperforms other methods at high noise. Yet, we verified that under exactly the same conditions as used in the present study, that method fails if $\sigma/\Delta_r > 0.6$. In other words: under otherwise similar experimental conditions, the present scheme would make it possible to detect steps that are nearly twice as small as those that can be resolved with existing method (for example: 5 nm instead of 8 nm). We should stress that the method that we have described thus far is, by construction, less flexible than the χ^2-method of ref. [6]. It is, however, more accurate: for instance, the χ^2-method typically yields a distribution of step sizes that may be rather wide. In one of the tests of ref. [6] at a noise level $\sigma/\Delta_r = 0.6$, steps of 8 nm were deemed to be correctly identified if the program yielded a step size between 5 and 11 nm. In contrast, for the highest noise levels that we studied, our estimate of the true step size is within ±3% of the true step size. This advantage of the present approach is arguably even more significant than the fact that we can detect steps at a 40% lower signal-to-noise ratio.

As explained above, we primarily compare the present approach with the technique of ref. [6]. There exist, of course, other methods that estimate step sizes without trying to reconstruct the underlying walk —for example the original method of ref. [1]. In appendix A we show that direct "histogram" methods already fail at a higher signal-to-noise ratio than the method presented here.

Thus far, we have assumed forward stepping only and we considered the case that the motors make steps of a single size. In contrast, the χ^2-method allows for steps of all possible sizes and signs. However, the present method can detect if its underlying assumptions are violated and, as we shall briefly discuss below, it can be extended to deal with more complex situations. In order to test if the present method can detect whether the underlying assumption of a single step size is justified, we generated an artificial data set with two step sizes. In units of the average step size $\langle\Delta\rangle$, the first step size (Δ_1) is equal to 0.625 and the second (Δ_2) is equal to 1.375. We take these two values as they correspond to the limiting cases that would be considered identical by the method of ref. [6]. If we perform the same analysis as before with this data set, but with a reference model that assumes a single step size of $\langle\Delta\rangle$, we observe that the free-energy modulations is incompatible with the data (see fig. 5). In other words: our approach tells us that these "experimental" data cannot be described by a model that assumes a single step size.

It is relatively straightforward to generalize the present method to the case where the motor can make steps of a fixed size both in the forward and the backward direction. With such an analysis we can determine not only the average step size but also the back-stepping probability. However, as the number of backward steps is typically rather small, the estimated error in the back-stepping probability is appreciable. It seems likely that this is a generic problem that is not limited to the present approach.

We have also explored the applicability of the present approach to walks that contain two step sizes. In such cases, the computational effort becomes appreciable. An example of such a calculation is shown in fig. 6. As can be seen from the figure, the absolute free-energy minimum is located close to the correct combination of step sizes. The figure suggests that there is a strong anti-correlation in the estimate of the two step sizes, in such a way that the average step size remains approximately constant. In view of the computational effort required, we did not attempt an analysis of the statistical accuracy of this calculation.

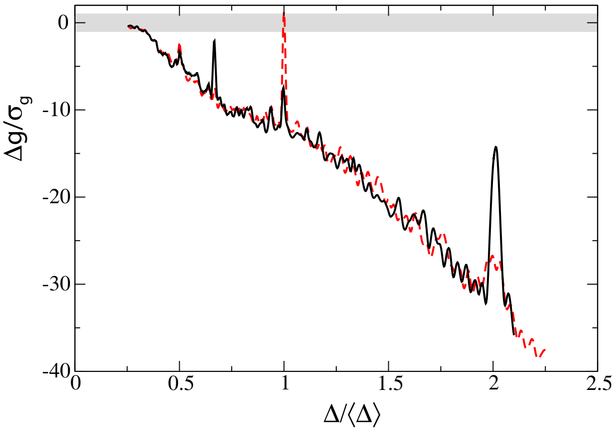

Fig. 5. (Color online) Distinguishing walks with one and two distinct step sizes. This figure compares the excess modulation amplitude of the free energy (see text) as obtained when the data set is generated using either a single step size Δ_r or with two step sizes ($0.625\Delta_r$ and $1.375\Delta_r$). The excess modulation is divided by σ', the estimated error in this quantity. Hence, the values ±1 correspond to deviations of $\pm\sigma'$. In both cases the free-energy modulation is computed assuming that the underlying walk has only a single step size. As can be seen from the figure, the single-step process (dashed curve) yields an excellent fit at $\Delta/\Delta_r = 1$. In contrast, walks with only a single step size cannot produce an adequate fit if the data result from a walk with two step sizes. In that case, the dominant peaks in the excess modulation amplitude are significantly different from zero. The gray band denotes the $\pm\sigma'$ error margin.

Finally, the present method works if the stepping process is Markovian. If this is not the case, one can still use "polymer-like" algorithms to detect both the steps size and the distribution of times between successive steps. However, at low signal-to-noise ratio, where the present approach performs best, such methods are not likely to yield meaningful information

Conclusion

In summary, we have presented a free-energy method to extract the step size and step frequency of processive molecular motors from noisy, experimental time position traces. For a simple forward walk with a single step size, the present method works well at noise levels that are inaccessible with existing techniques. Moreover, it provides very accurate estimates of the step size. The method allows us to detect if multiple step sizes or back-stepping is important. However, whatever the analysis method, such probability of back-stepping can only be estimated with limited accuracy for time traces of a realistic length.

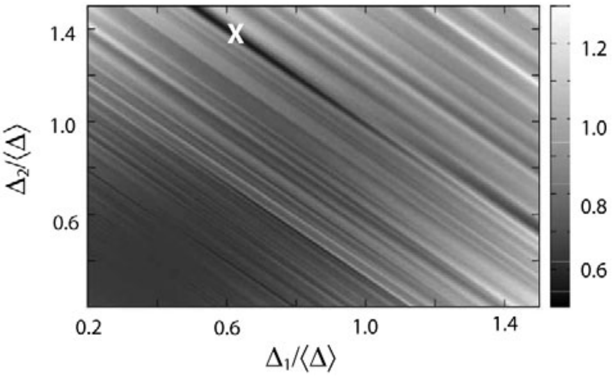

Fig. 6. Analysis of same data as in fig. 5 but now assuming two distinct step sizes. The figure shows the free energy as a function of the two step sizes Δ_1 and Δ_2. The figure is not symmetric under permutation of Δ_1 and Δ_2 because the first step is constrained to have a size Δ_1. The gray scale denotes the value of the free energy: dark gray corresponds to low free energy. The white cross denotes the true combination $\{\Delta_1, \Delta_2\}$. As can be seen, this point is close to the free-energy minimum. However, this minimum is very shallow along the line $\Delta_1 + \Delta_2 = 2$.

Finally, we note that the present method is quite general and could be applied to a variety of other stochastic processes where many discrete "histories" are compatible with noisy experimental data.

We are grateful to S. Tanase-Nicola, J. Husson, P. Shaklee, B. Mulder, K. Visscher and M. Dogterom for numerous illuminating discussions, J. Husson and M. Dogterom for sharing with us their experimental results and numerical code of ref. [6], and K. Visscher for sharing with us his results and numerical code prior to publication [7]. The work of the FOM Institute is part of the research program of FOM and is made possible by financial support from the Netherlands organization for Scientific Research (NWO). SJC and DF acknowledge financial support from the Royal Society of London (Wolfson Merit Award) and DF acknowledges support from the ERC (Advanced Grant agreement 227758).

Appendix A. Histogram methods

Consider a data set $\{X(t)\}$ of motor displacements *versus* time. To make a histogram, we project the data points on the ordinate axis ("y-axis"). For convenience, we assume that there are as many bins in the vertical direction as data points (say N). This is not essential. If the steps were perfectly sharp (no noise), we would find equally spaced δ-like peaks in this histogram with a spacing Δ_r and zeros in between. With noise, the δ-functions are broadened to Gaussians with a width σ. The resulting histogram is then a convolution of the δ-comb and the (normalised) Gaussians.

Fourier transforming this signal only yields a signal at $\omega = n2\pi/\Delta_r$. The dominant peak is the one with $n = 1$. The amplitude S (stands for "Signal") is given by

$$S = Ne^{-\omega^2\sigma^2/2} = Ne^{-2\pi^2(\sigma/\Delta_r)^2}.$$

To estimate the noise, we use the fact that we can get the power spectrum of the noise from the Fourier transform of the noise auto-correlation function. Assuming that the signal is not large compared to the noise (otherwise we need not bother), and that the noise is δ-correlated, we expect that the root-mean-square noise amplitude is $\sqrt{N}$.

Then the signal-to-noise ratio should be given by

$$\frac{S}{N} \approx N^{1/2}e^{-2\pi^2(\sigma/\Delta_r)^2}.$$

Now consider a realistic "noisy" case: $N = 4000$, $\Delta_r/\sigma = 1$, *i.e.* at the limit of the method described in the present paper. Then

$$\frac{S}{N} \approx 4000^{1/2}e^{-2\pi^2} = 10^{-7}.$$

In other words: the steps would be undetectable.

In order to get $S/N \approx 1$, we need $\sigma/\Delta_r \approx 0.45$. The method of Svoboda *et al.* is slightly different. These authors construct a histogram of distances between successive points. For a single-step process, this histogram would show two peaks: one at 0 and one at Δ_r. Both are broadened by a noise $\sigma\sqrt{2}$. The amplitude of the peak at Δ_r is N_{steps}/N times weaker than the peak at zero. To distinguish the signal from the noise we need

$$N_{\text{steps}} \geq Ne^{-\Delta_r^2/4\sigma^2}.$$

If we consider the case $N_{\text{steps}}/N = 1/100$, then $\sigma/\Delta_r \leq 0.23$ which is worse than the Fourier-transform histogram method.

References

1. K. Svoboda, C.F. Schmidt, B.J. Schnapp, S.M. Block, Nature **365**, 721 (2008).
2. N.J. Carter, R.A. Cross, Nature **435**, 308 (2005).
3. S. Toba, T.M. Watanabe, L. Yamaguchi-Okimoto, Y.Y. Toyoshima, H. Higuchi, Proc. Natl. Acad. Sci. U.S.A. **103**, 5741 (2006).
4. H. Yardimci, M. van Duffelen, Y. Mao, S.S. Rosenfeld, P.R. Selvin, Proc. Natl. Acad. Sci. U.S.A. **105**, 6016 (2008).
5. B.C. Carter, M. Vershinin, S.P. Gross, Biophys. J. **94**, 306 (2008).
6. J.W. Kerssemakers, E.L. Munteanu, L. Laan, T.L. Noetzel, M.E. Janson, M. Dogterom, Nature **442**, 709 (2006).
7. B. Kalafut, K. Visscher, Comput. Phys. Commun. **179**, 716 (2008).
8. W. Hua, E.C. Young, M.L. Fleming, J. Gelles, Nature **388**, 390 (1997).

9. C.J. Lawrence, R.K. Dawe, K.R. Christie, D.W. Cleveland, S.C. Dawson, S.A. Endow, L.S.B. Goldstein, H.V. Goodson, N. Hirokawa, J. Howard, R.L. Malmberg, J.R. McIntosh, H. Miki, T.J. Mitchison, Y. Okada, A.S.N. Reddy, W.M. Saxton, M. Schliwa, J.M. Scholey, R.D. Vale, C.E. Walczak, L. Wordeman, J. Cell Biol. **167**, 19 (2004).
10. M. Rief, R.S. Rock, A.D. Mehta, M.S. Mooseker, R.E. Cheney, J.A. Spudich, Proc. Natl. Acad. Sci. U.S.A. **97**, 9482 (2000).
11. David J.C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, 2003).
12. G.C.A.M. Mooij, D. Frenkel, Mol. Phys. **74**, 41 (1991).
13. B. Bozorgui, D. Frenkel, Phys. Rev. E **75**, 036708 (2007).